USCMS Engineer Status Report for April 2004

Bill Tanenbaum

May 10, 2004

1 Work Performed This Month

1) Wrote script to recreate the metadata needed for the DCO4 JetMet analysis. Most of the effort was in attaching the runs to the metadata, which involved invoking findColls and AttachRun. This script was an ongoing effort, and was updated and improved several times throughout the month.

2) In conjunction with the metadata script, I developed several improvements to COBRA to aid in the reconstruction of missing or incomplete metadata:

a) Allowing multiple runs to be attached to the metadata in one operation.

b) Improving the attach run code to simultaneously fix the collection information, eliminating the need for separately running FixColl. This is important because FixColl opens all the event files in a dataset, and thus cannot be used for large datasets in dcache.

c) Wrote a new command, SortColl, which sorts the attached runs if they were not attached inthe correct order. This command is fast because it accesses only the metadata.

d) Improving the attachRun code to find the collection information automatically, to eliminate the need for running findColls. (Note: this has not yet been checked into CVS, because it is stillbeing improved. However, it works.)

3) Months ago, I rewrote the ROOT plugin (part of ROOT) for dcache in order for it to be able to be used with POOL. Therefore, I now seem to own it. When CDF tested it, they discovered performance problems. On opening a file, dcache was accessed unnecessarily multiple times. I optimized the plug-in to eliminate the unnecessary accesses, and submitted the updates to ROOT. NOTE: There was no performance problem for CMS because we

locally mount the dcache file system in /pnfs. The performance problem only affected users who access dcache that is not locally mounted.

2 Plans For Next Month

On demand update of missing or incomplete metadata as needed by the analysis, thereby eliminating the need for advance reconstruction of the metadata.